

PT-003 CON
TITLE OF THE INVENTION

EFFICIENT TRANSMISSION OF DATA TO MULTIPLE NETWORK NODES
CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims priority to and is a continuation of U.S. Application Serial No. 09/248,371, entitled "Efficient transmission of data to multiple network nodes" filed on February 8, 1999.

BACKGROUND OF THE INVENTION

The growth of the Internet has spurred the development of many new programs and services that permit two-way communication between more than two users at the same time. Such communication may be in the form of typed messages, for example in the internet relay chat (IRC) service, or as voice or video transmitted over the network.

Typically such programs and services use the concept of a "channel" to refer to a set of network nodes that are in mutual communication with each other. When a node subscribes to a channel, it receives copies of all of the messages that are transmitted over the channel. Programs and services using channels must therefore, provide a mechanism for distributing packets issuing from one node on the channel to all of the other nodes on the channel.

Two general models have been heretofore applied in this regard. In a peer-to-peer model each node is responsible for sending a copy of each packet to all of the other nodes on the channel. This model is limited, however, by the bandwidth of the node's network connection. For example, if the node was connected to a network by a slow dial-up connection and the node were to transmit to more than a few other nodes on the network, this could produce excessive delays. Delays are especially significant for real

PT-003 CON

time audio or video communication, since they can reduce the quality of the audio or video signal received by other nodes.

An alternative to the peer-to-peer model is the client-server model. In the client server model, data communicated between the various nodes subscribing to a channel (referred to as "clients") are always directed through a dedicated intermediary computer known as the "server." In the simplest case, the server acts as a reflector of the data packets received from clients. Each packet received at the server is copied, with a copy transmitted to every other client on the channel.

Alternatively, servers may combine data received from multiple nodes (e.g., by mixing simultaneously received audio signals), and transmit the combined data to the clients on the channel.

Because the server is dedicated to retransmission of data from a large number of clients, its bandwidth may be underutilized if it is not operating at full capacity.

SUMMARY OF THE INVENTION

The present invention provides an improved method, system, and computer program product for providing communication among multiple nodes on a network.

In one aspect, the invention is a method for transmitting a data block over a network from a first sending node to a first set of recipient nodes, including, in the first sending node, a) dividing the first set of recipient nodes into a subset of selected nodes, selected according to scoring criteria associated with each recipient node, and a subset of unselected nodes, b) assigning at least one of the unselected nodes to at least one selected node according to scoring criteria associated with the respective selected nodes, and c)

PT-003 CON

transmitting to each selected node a packet including the data block and a list of the nodes assigned to the selected node.

In another aspect, the invention is a method for transmitting a data block over a network from a first sending node to a first set of recipient nodes, including, the steps of, in a selected node, a) receiving from the first sending node the packet including the data block and a list of assigned nodes, b) dividing the list of assigned nodes into a subset of selected assigned nodes, selected according to scoring criteria associated with each assigned node, and a subset of unselected assigned nodes, c) reassigning at least one of the unselected assigned nodes to at least one selected assigned node according to the scoring criteria associated with respective selected assigned nodes, and d) transmitting to each selected assigned node a packet including the data block and a list of the nodes re-assigned to the selected assigned node.

It is therefore an advantage of the invention to provide a method and system for enabling a node in a data network, having a low effective bandwidth, to transmit data to a large number of other nodes in the network, without excessive delay. It is a further advantage of the invention to provide a method and system for enabling a node in a data network to transmit data to a large number of other nodes in the network, without the use of a dedicated server.

These, and other advantages of the present invention, are fully described in the following detailed description of a preferred embodiment, and in the claims.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows a network for performing an embodiment of the invention.

PT-003 CON

Figure 2 is a schematic diagram of a packet for transmitting data on a network.

Figure 3 is a schematic diagram showing channel lists in a node.

Figure 4 is a flowchart showing steps taken by an embodiment of the invention.

Figure 5 is a schematic diagram of a packet used by an embodiment of the invention.

Figure 6 is a flowchart showing steps taken by an embodiment of the invention.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

Figure 1 shows a computer network 100, such as the Internet, or a local area network, in which multiple nodes 110a-110h are connected to the network using a set of heterogeneous communication interfaces 120a-120h. Examples of such communication interfaces include 28.8 kilobaud modems, ISDN lines, T1 lines and T3 lines. Each communication interface 120a-120h has a characteristic bandwidth, measured in bits per second: A node may include any programmed electronic device capable of communicating data over a network, such as a computer workstation, a cable television interface, a personal communications system (PCS) device, and the like.

Nodes communicate by transmitting packets 200 via the network. As shown in Figure 2, each packet includes a header 210, that contains addressing information, such as an internet protocol (IP) address, and a payload 220 that contains the data to be communicated to the recipient.

PT-003 CON

Referring to Figure 3, each node on the network may, at a given time, be subscribed to one or more channels. Each node 110 maintains a channel list for each channel to which it is subscribed. A channel list 300 includes a channel ID 310, which uniquely identifies the channel, along with a set of node IDs 320, which identify every other node subscribed to the channel. Associated with each node address in the channel list 300 is an effective bandwidth value 330, a number representing the effective bandwidth of the communications interface for that node.

Channel subscription (i.e., the adding or subtracting of nodes from a channel) may be handled using any conventional method. For example, a dedicated subscription server may be used to subscribe and unsubscribe nodes on a channel. A new node could be added to the channel by transmitting a request to the server. The server, in turn, would inform all existing channel subscribers of the identity of the new member, which would update their respective channel lists 300 accordingly. Alternatively, one of the channel members could be selected to serve as the subscription server through a negotiation process between channel members.

When a node subscribes to a channel, it provides the subscription server with information about itself, including its effective bandwidth, which is retransmitted to the other channel subscribers to permit each subscribing node 110 to maintain an up-to-date channel list 300.

In general, the effective bandwidth of a node is the nominal bandwidth of the node's communications interface. However, under some circumstances, a node may provide an effective bandwidth value that is less than the nominal bandwidth of its communications interface when it cannot allocate all of its bandwidth for communication

PT-003 CON

on a given channel. For example, the node might already be subscribed to other communication channels, or the node may be engaged in non-channel communication.

The transmission of a data packet from a sending node to other nodes in a channel will now be described. First, the sending node 110 determines the value of MAXCHAN, the maximum number of simultaneous transmission streams that it can support (step 500). MAXCHAN may be a function of the node's effective bandwidth, or may be empirically determined.

Next, the sending node 110 examines the channel list 300 to determine whether there are more than MAXCHAN nodes on the channel (other than the sending node) (step 510). If there are not more than MAXCHAN other nodes on the channel list 300, the sending node sends packets containing the data block directly to each of the nodes on the channel list (step 520). In a preferred embodiment, the packets are sent approximately simultaneously.

If there are more than MAXCHAN other nodes on the channel list, the sending node selects the MAXCHAN number of nodes on the channel list having the highest effective bandwidths (step 530) .

Next, the sending node computes, for each of the selected nodes, a fair share value (step 540). This value is computed by taking the ratio of the effective bandwidth 330 for a given selected node to the sum of the effective bandwidths 330 for all of the selected nodes.

Next, the sending node assigns each of the remaining (unselected) nodes to one of the selected nodes, in approximate proportion to the fair share values computed for each selected node (step 550). In an alternative embodiment, the sending node could

PT-003 CON

assign only a portion of the unselected nodes, according to a prioritization scheme, if the node determines that the number of unselected nodes exceeds an operational limit.

Next, the sending node sends a packet to each of the selected nodes (step 560). In a preferred embodiment, the packets are sent approximately simultaneously. Referring to Figure 5, the header 210 of each packet is addressed to one of the selected nodes, while the payload portion 220 of each packet includes node ID of the sending node (identified as the originator of the data block) 225 the data block 230, and a list 240 of the nodes assigned to the node to which the packet is addressed.

Referring now to Figure 6, when a packet from a sending node arrives at one of the selected nodes, it is received (step 600), and the originator node ID 225 and data block 230 are extracted (step 610). This information can then be used at the receiving node by other processes. For example, if the data block 230 includes audio data, the data block 230 could be used to generate audio outputs at the selected node.

The receiving node then determines its MAXCHAN value, (step 620), and compares it to the number of nodes on the list of assigned nodes in the packet received from the sending node (step 630). If the number of nodes in the list is not greater than MAXCHAN, the receiving node transmits packets containing the data block 230 to each of the assigned nodes (step 640).

If the list of assigned nodes is longer than MAXCHAN, the following steps are performed. The receiving node selects the MAXCHAN nodes from the assigned nodes list having the greatest bandwidth (step 650), and then calculates a fair share value for each of the selected nodes (step 660). It then re-assigns each of the unselected assigned nodes from the list to one of the selected nodes from the list, in approximate proportion to the fair

PT-003 CON

share values of each selected node (step 670). (In an alternative embodiment, the receiving node could re-assign only a portion of the unselected assigned nodes, according to a prioritization scheme, if the receiving node determines that the number of unselected assigned nodes exceeds an operational limit.) The receiving node then transmits a packet to each of the selected nodes, containing the originator node ID 225, the data block 230, plus a list of the nodes re-assigned to the node to which the packet is addressed (step 680). Each of the nodes receiving these packets processes them in a similar fashion, by executing the steps beginning at step 600. This process continues until the data block has been distributed to all of the nodes in the channel.

In this way, the invention advantageously provides method for distributing packets to a large number of recipient nodes on a channel, without effective limitation from the effective bandwidth of the sending node, and without using a dedicated server.

In alternative embodiments, criteria other than, or in addition to, effective bandwidth, may be used by a sending node to select recipient nodes.

In one alternative embodiment, the channel list 300 of a given node includes information identifying the latency (average delay between transmission and reception of a packet) associated with transmissions from the given node to each other node in the channel. These values are computed by the given node and updated periodically by sending test messages to the other nodes on the channel and calculating the time to receive a response. In this embodiment, nodes are selected by assigning to each a score, based on a weighted difference of the effective bandwidth and the latency of the node, and then selecting the nodes with the highest scores. Similarly, unselected nodes are assigned to

PT-003 CON

selected nodes in approximate proportion to their scores, rather than their effective bandwidths.

In another embodiment, node scores are computed as a function of the effective bandwidth, latency, and recency. Recency refers to the amount of time since a data block was received that was initially sent (i.e., not merely retransmitted as described above) by a given node. Such a function could give greater scores to nodes with lower recency values, thus making it more likely that nodes from which packets were recently received will be among the first to receive outgoing data.

In another alternative embodiment, a given unselected node may be assigned to more than one of the selected nodes, thus providing a redundant route between the sending node and the ultimate recipient for transmission of a given data block. If this resulted in multiple copies of the data block arriving at the ultimate recipient node, the ultimate recipient node would simply ignore the subsequent copies. This could be done, for example, by assigning each data block a serial number in the sending node, and comparing the serial number of each data block received by the ultimate recipient node to a list of serial numbers already received from the sending node. This embodiment would be useful where the reliability of different transmission paths in the network varies or is unknown.

The invention described herein may be implemented in software or in hardware. In particular, the invention may be implemented by programming a computer in each node to perform the steps necessary to carry out the method described herein. Programming codes for carrying such steps may be placed on a computer readable medium for convenient transport and storage.